Equivalence of Probabilistic Tournament and Polynomial Ranking Selection

Kassel Hingee – Department of Mathematics @ Australian National University – `hinkass@gmail.com`

Marcus Hutter – RSISE @ ANU and SML @ NICTA, Canberra, ACT, 0200, Australia – `marcus@hutter1.net`

March 2008

Abstract

Crucial to an Evolutionary Algorithm's performance is its selection scheme. We mathematically investigate the relation between polynomial rank and probabilistic tournament methods which are (respectively) generalisations of the popular linear ranking and tournament selection schemes. We show that every probabilistic tournament is equivalent to a unique polynomial rank scheme. In fact, we derived explicit operators for translating between these two types of selection. Of particular importance is that most linear and most practical quadratic rank schemes are probabilistic tournaments.

1 Introduction

Evolutionary algorithms. Evolutionary Algorithms (EAs) are probabilistic search algorithms based on evolution [Gol89, ES03]. They operate by exploiting the information contained in a population of possible solutions (via similarities between individuals). The aim is to find an individual that maximises (or minimises) an objective function, which maps from individuals to the real line. The population is transformed by first selecting individuals. Mutation and/or recombination is then used to either replace a few individuals from the population or create an entirely new population.

Standard selection methods. The most prevalent methods for selecting individuals are proportionate, linear rank, tournament, and truncation [HL06]. In proportionate selection individuals are chosen with a probability proportional to their fitness (the value of the objective function evaluated at the individual) [Bäc94]. A common method to gain more control over selection pressure, is to scale the fitness values before the selection is made [Bäc94]. Linear ranking proceeds by ordering the population according to their fitness. The chance that an individual is selected is then a linear function of its (unique) rank [Bäc94]. Tournament selection creates a tournament by randomly choosing t individuals, the best individual in the

tournament is then selected. For truncation selection the k fittest individuals have uniform probability of selection, while the remainder have zero chance of being selected.

The choice of selection scheme is crucial to algorithm performance. If the selection pressure is too high then diversity of the population decreases rapidly and the algorithm converges prematurely to local optima or worse. With too little pressure there is not enough push toward better individuals and the population takes too long to converge. Many methods to choose or adapt the selection pressure or avoid the problem otherwise have been invented (see [HL06] for some references). A particularly simple one is fitness uniform selection, which uniformly selects a fitness value, and then the individual with fitness closest to this value.

It is quite profitable to study selection schemes due to their generality. They depend only on the set of fitness values and not on the rest of the algorithm. Hence their behaviour can be studied in isolation and the results applied to any evolutionary algorithm. In this paper we introduce and study generalizations of rank and tournament selection (both actually only depend on the rank and not the absolute fitness value itself).

Polynomial rank selection. Linear ranking has a small range of selection pressures (from [Bäc94], for a population of n individuals the probability that the fittest individual is selected must be between $1/n$ and $2/n$), but it has the flexibility of a real-valued parameter that can vary continuously (the slope of the linear function). Ranking schemes with high selection pressures, such as when the probability of selection is an exponential function of the rank, have occasionally been used [WC02]. It is natural then to generalise from linear to polynomial functions to cover the instances where medium pressure is required. Hence the probability of an individual with rank k being selected with a polynomial rank scheme of degree d is:

$$P(I = k) = \sum_{l=1}^{d+1} a_l k^{l-1} \qquad (1)$$

where $a_l \in \mathbb{R}$ are parameters defined by the algorithm designer. For simplicity we assume that selection is per-

formed with replacement and each individual has unique rank, however our results still hold when there are ties in the rank. The only restriction on the a_k is that they must produce a proper probability distribution, i.e. for a population of n individuals: $P(I=k) \geq 0$ for all $k=1,2,...,n$ and $\sum_{k=1}^{n} P(I=k)=1$. Hence, while the population is ordered, the schemes may favour low ranks, high ranks or neither, depending on the choice of the (a_l).

This selection method encompasses the low pressures of linear schemes ($a_3=...=a_{d+1}=0$) and can give good approximations of the high pressure exponential cases (via Taylor polynomials). Furthermore the wealth of general knowledge about polynomials means that while it has numerous parameters (coefficients of the monomials), it is also easy to predict their impact.

Probabilistic tournament selection. Tournament selection has a large range [Bäc94], but a discrete parameter, leaving the possible selection pressures somewhat restricted. This can be overcome by selecting probabilistically from the tournament, rather than always choosing the best in the tournament. However the extra parameters required are not easy to understand. Their precise effect on the behaviour is not at all obvious. Probabilistic Tournament selection still only sorts $t \ll n$ individuals, making it much faster than any ranking scheme.

Let i_s be the (rank of the) individual in position $s \in \{1,...,t\}$ of the rank-ordered tournament. We call s the *seed* of i. Let $P(I_s=k)$ be the probability that seed s has rank k. In any given tournament, the probability that the seed s individual is chosen will be a user defined constant α_s. Then the probability of an individual k being selected through a size t probabilistic tournament is:

$$P(I = k) = \sum_{s=1}^{t} \alpha_s P(I_s = k) \qquad (2)$$

Standard (deterministic) tournament always selects the individual of highest rank in the tournament, i.e. $\alpha_1 = 1$ and $\alpha_2 = .. = \alpha_t = 0$.

To ensure that choosing a winner from the tournament makes sense, the α_s must satisfy the probability constraints $\alpha_s \geq 0 \, \forall s$ and $\sum_{s=1}^{t} \alpha_s = 1$. We assume that the tournament is created by random selection *with* replacement and for now that each individual in the population has a unique fitness. This defines $P(I_s=k)$ (Section 2). Note that even if every individual in the population is unique, it is possible for it to be repeated in the tournament.

Previous work on the relation between rank and tournament selection. In this paper we investigate the equivalence between the generalised schemes (1) and (2) with the aim of providing a scheme that combines the superior understanding of polynomial rank with the speed of probabilistic tournament.

Bäck [Bäc94] found that an individual's chance of selection in deterministic tournament selection is a polynomial, hence each is equivalent to a polynomial rank selection method. Wieczorek and Czech [WC02], and Blickle [BT95] arrived at the same conclusion using a different method. So while the name 'polynomial rank selection' is new, its concept is fairly old.

The study of probabilistic tournaments isn't new either: Hutter [Hut91, p.11] proved that every size 2 probabilistic tournament is a linear rank scheme, and Goldberg [GD91] did the same but only for a continuous population. Fogel [Fog88] applied to the traveling salesman problem, a variation wherein each individual underwent numerous t=2 tournaments. The probability of winning each tournament was dependent on the fitness of the individuals involved and the individuals selected were those with the highest number of wins.

Contents: Equivalence of polynomial rank and probabilistic tournament selection. We extend these results by finding that every t sized probabilistic tournament is equivalent to a polynomial rank scheme with a polynomial degree of $d=t-1$ or less (Section 2). We continue on to show that the equivalence is unique (Section 3), and give an explicit expression for the inverse map (Section 4). This allows the establishment of simple criteria for polynomial rank schemes that are probabilistic tournaments (Section 5). Unfortunately not every possible polynomial rank scheme satisfies the criteria, but most (and in the limit of an infinite population, all) linear and most "interesting" quadratic ones are equivalent to probabilistic tournaments. This is good enough for all practical purposes, if it generalises to higher order polynomials.

Notation. Throughout the paper we use the following notation. If not otherwise indicated, an index has the full range as defined in this table.

Symbol	Explanation
δ_{ij}	Kronecker symbol ($\delta_{ij}=1$ for $i=j$ and $\delta_{ij}=0$ for $i \neq j$)
n	Number of individuals in the population
i,j,k	Rank (unique label) of individuals $\in \{1,...,n\}$
ι, κ	Rank indices that only run from $1,...,t$
p,q,r,s	Seed index $\in \{1,...,t\}$
I_s	Rank of the individual with seed s
I	Rank of the individual selected
$\pi_i = P(I=i)$	Probability that i is selected
l	Polynomial coefficients index $\in \{1,...,d+1\}$
a_l	Coefficients of x^{l-1} for the polynomial
$\alpha_1, \alpha_2,...,\alpha_t$	Tournament selection coefficients
$x_i, \boldsymbol{x}, (x_i)$	Vector $\boldsymbol{x} = (x_i) = (x_1,...,x_t)$
Δ_m	$m-1$ dimensional probability simplex

2 Probability of Selection via a Tournament

In this section we find the probability of an individual being successful (the winner) via tournament selection. This

will provide a formula for an equivalent ranking selection scheme. It is sufficient to consider just one selection event in isolation, since we consider selection with replacement.

We assume a population $\mathcal{P}$ consisting of n individuals $c_1,...,c_n$ with fitness $f_1,...,f_n$. Without loss of generality we assume that they are ordered, i.e. $f(i) \geq f(j)$ for all $j \leq i$. For now we also assume that all fitness values are different, hence individual c_i has rank i. The rank is all we need in the following, and we will say "individual i", meaning "individual c_i".

Definition 1 (polynomial rank selection)
Polynomial $\boldsymbol{a}$-ranking selects individual c_k from population $\mathcal{P}$ with probability $P(I\!=\!k)\!=\!\sum_{l=1}^{d+1} a_l k^{l-1}$

Definition 2 (probabilistic tournament selection)
A probabilistic $\boldsymbol{\alpha}$-tournament selects t individuals from population $\mathcal{P}$ uniformly at random with replacement. Let c_{I_s} be the individual of rank s in the tournament, called seed s (while it has rank I_s in the population). Finally the seed s individual, I_s, is chosen with probability $\alpha_s\!=\!P(I\!=\!I_s)$ as the winner I.

Theorem 3 (tournament=polynomial)
Probabilistic $\boldsymbol{\alpha}$-tournament selection coincides with polynomial $\boldsymbol{a}$-ranking (for $d\!=\!t\!-\!1$ and suitable $\boldsymbol{a}$).

Proof. We derive an explicit expression for the probability π_k that the tournament winner has rank k. Any seed s may have rank k ($I_s\!=\!k$) and may be the winner ($I\!=\!I_s$), hence

$$\pi_k \;\equiv\; P(I=k) \;=\; \sum_{s=1}^{t} P(I=I_s)P(I_s=k) \;=\; (2)$$

where we have exploited that by definition the probability that $I=I_s$ is independent of the rank $I_s=k$. $P(I_s=k)$ is the probability that seed s has rank k. It is difficult to formally derive an expression for $P(I_s=k)$, but we can easily get it by considering distribution functions. The probability of an individual selected into the tournament having a particular rank is $1/n$, hence having rank equal to or less than k is k/n and larger than k is $1-k/n$. Further, $I_r \leq k \wedge I_{r+1} > k$ if and only if r seeds have rank $\leq k$ and $t-r$ seeds have rank $> k$, hence

$$P(I_r \leq k \wedge I_{r+1} > k) \;=\; \binom{t}{r}(\tfrac{k}{n})^r(1-\tfrac{k}{n})^{t-r}$$

since there are $\binom{t}{r}$ ways of choosing r individuals with rank $\leq k$ from t individuals. The above expression is a polynomial in k of degree t. Together with

$$P(I_s \leq k) \;=\; \sum_{r=s}^{t} P(I_r \leq k \wedge I_{r+1} > k),$$

we get the explicit expression

$$P(I_s = k) \;=\; P(I_s \leq k) - P(I_s \leq k-1) \qquad (3)$$
$$=\; \sum_{r=s}^{t} \binom{t}{r}\big[(\tfrac{k}{n})^r(1-\tfrac{k}{n})^{t-r} - (\tfrac{k-1}{n})^r(1-\tfrac{k-1}{n})^{t-r}\big]$$

Using the binomial theorem to find the k^t and k^{t-1} coefficients in the square brackets above reveals that the former coefficients cancel out while the latter do not. This implies that $P(I_s\!=\!k)$ is a polynomial in k of degree (at most) $t-1$, and thus the weighted average (2) is as well. Summing (2) over the population yields $\sum_{k=1}^{n} P(I\!=\!k)\!=\!1$, as it should, since the tournament coefficients are such that some individual is always chosen. Consequently, every tournament is a polynomial rank scheme of degree at most $t-1$ (one can choose $\boldsymbol{\alpha}$ such that it is of lower degree). $\qquad\square$

Examples. Expression (3) can be rewritten as

$$P(I_s = k) \;=\; \sum_{r=0}^{s-1} \binom{t}{r}\big[(\tfrac{k-1}{n})^r(1-\tfrac{k-1}{n})^{t-r} - (\tfrac{k}{n})^r(1-\tfrac{k}{n})^{t-r}\big]$$

which will be convenient in the following examples. Standard tournament always selects I_1 ($\alpha_1\!=\!1$), hence [Bäc94]

$$P(I = k) \;=\; P(I_1 = k) \;=\; (1 - \tfrac{k-1}{n})^t - (1 - \tfrac{k}{n})^t$$

See Figure 1. For $t=1$ there is no selection pressure, $P(I\!=\!k)\!=\!\tfrac{1}{n}$. For $t=2$ we get

$$P(I_1 = k) = \tfrac{2n-2k+1}{n^2} \quad \text{and} \quad P(I_2 = k) = \tfrac{2k-1}{n^2}$$

Hence probabilistic tournaments of size 2 lead to linear ranking [Hut91]

$$P(I=k) \;=\; \alpha_1 P(I_1=k) + \alpha_2 P(I_2=k) \;=\; a_1 + a_2 k,$$
$$a_1 = \tfrac{1}{n^2}[(2n+1)\alpha_1 - \alpha_2], \qquad a_2 \;=\; \tfrac{2}{n^2}(\alpha_2 - \alpha_1) \quad (4)$$

Remark. More interesting is actually the converse, replacing rank selections by equivalent efficient tournaments. Before we can answer this, we need to break down (3) into a product of simple regular matrices.

3 The Map from Tournament to Polynomial is Unique

The next natural question is whether different tournament bias $\boldsymbol{\alpha}$ implies different selection probability. It seems plausible that the maps from tournaments $\boldsymbol{\alpha}$ to rank probabilities $\boldsymbol{\pi}$ and to polynomial coefficients $\boldsymbol{a}$ are injective, but the proof is fairly involved. The good news is that construction in the proof allows us to find a closed form expression for the desired inverse. Let $\Delta_m = \{\boldsymbol{x} \in \mathbb{R}^m : x_i \geq 0\, \forall i, \sum_{i=1}^{m} x_i = 1\}$ be the $m-1$ dimensional probability simplex, i.e. $\boldsymbol{\pi} \in \Delta_n$ and $\boldsymbol{\alpha} \in \Delta_t$.

Theorem 4 (tournament→polynomial)
The function $R:\Delta_t \to \Delta_n$ in (2), mapping tournament probabilities $\boldsymbol{\alpha}$ to rank probabilities $\boldsymbol{\pi}$, is total, linear, and injective:

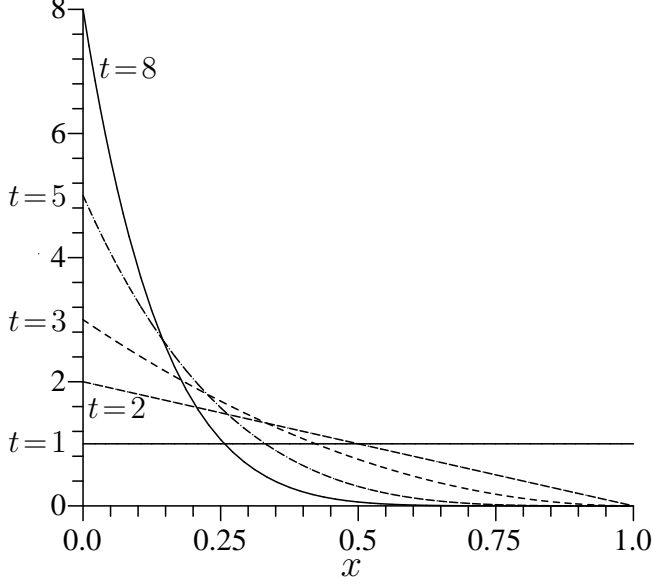$$\pi_k = P(I=k) = \sum_{s=1}^{t} R_k^s \alpha_s, \quad i.e. \quad \boldsymbol{\pi} = R\boldsymbol{\alpha},$$

Figure 1: [tournament probabilities for large n] *Probability density $nP(I_1 = xn)$ that the tournament winner has rank xn, for tournament size $t = 1, 2, 3, 5, 8$.*

where $R_k^s = P(I_s = k)$ is defined in (3). Matrix R can also be written as a product $R = \overline{D}PFCD = VNFCD = VT$ with matrices $\overline{D}$, P, F, C, D, V, and N defined in (7), (9), (10), (8), (6), (12), and (13). Similarly, the function $T : \Delta_t \to \mathbb{R}^t$, mapping $\boldsymbol{\alpha}$ to polynomial coefficients $\boldsymbol{a}$, is unique, linear and injective:

$$a_l = \sum_{s=1}^{t} T_l^s \alpha_s, \quad i.e. \quad \boldsymbol{a} = T\boldsymbol{\alpha},$$

where matrix $T = NFCD$.

Proof. Tournament always selects one individual from $\mathcal{P}$ as the winner, hence $R\boldsymbol{\alpha} \in \Delta_n$ for every $\boldsymbol{\alpha} \in \Delta_t$. See the proof of Theorem 3 for how to prove this formally.

Matrices H and G. We now prove injectivity. With

$$H_k^r := (\tfrac{k}{n})^r (1 - \tfrac{k}{n})^{t-r} \quad \text{and} \quad G_k^r := \binom{t}{r}(H_k^r - H_{k-1}^r)$$

we can write (3) as

$$R_k^s \equiv P(I_s = k) = \sum_{r=s}^{t} G_k^r \qquad (5)$$

Einstein notation. Einstein's sum convention will be convenient in the following argument: When an index occurs repeatedly in the multiplication of two objects, a sum over the index over its full range is implicitly understood, e.g. $G_k^r D_r^s$ means $\sum_{r=1}^{t} G_k^r D_r^s$.

Lower-triangular matrix D. The lower-triangular matrix

$$D_r^s := \begin{cases} 1 & \text{if} \quad s \leq r \\ 0 & \text{if} \quad s > r \end{cases} \qquad (6)$$

has the property that $\sum_{r=1}^{t} G_k^r D_r^s = \sum_{r=s}^{t} G_k^r$. Using Einstein's sum convention this allows us to rewrite (5) as

$$R_k^s = G_k^r D_r^s$$

i.e. as a product of an $n \times t$ matrix G with a $t \times t$ matrix D.

Inverse of D. The "inverse" of D is:

$$\overline{D}_k^i := \left\{ \begin{array}{cl} 1 & \text{if } k = i \\ -1 & \text{if } i = k - 1 \\ 0 & \text{otherwise} \end{array} \right\} = \delta_{k,i} - \delta_{k-1,i} \qquad (7)$$

This is a matrix with 1 on the primary diagonal; -1 on the diagonal that is below the primary diagonal; and 0 otherwise.

Decomposing G. G_k^r itself can actually be decomposed into $\overline{D}_k^i$ and H_i^q and a pure diagonal matrix

$$C_q^r = \binom{t}{q}\delta_{q,r} \qquad (8)$$

comprised of the binomial coefficients:

$$G_k^r = (H_k^q - H_{k-1}^q)C_q^r = \overline{D}_k^i H_i^q C_q^r$$

(note that $\overline{D}$ is the inverse of an $n \times n$ sized D matrix here).

Decomposing H into P and F. We can decompose H_i^q further be using the binomial identity:

$$\begin{aligned} H_i^q &= (\tfrac{i}{n})^q (1 - \tfrac{i}{n})^{t-q} \\ &= (\tfrac{i}{n})^q \sum_{s=1}^{t-q} \binom{t-q}{s}(\tfrac{-i}{n})^{t-q-s} \\ &= \sum_{s=1}^{t-q} (-)^{t-q-s}\binom{t-q}{s}(\tfrac{i}{n})^{t-s} \\ &= \sum_{p=q}^{t} (-)^{p-q}\binom{t-q}{t-p}(\tfrac{i}{n})^p \end{aligned}$$

So $H_i^q = P_i^p F_p^q$, where P is a matrix of monomials:

$$P_i^p := (\tfrac{i}{n})^p, \qquad (9)$$

and F is a lower-triangular matrix composed of various binomials:

$$F_p^q := \left\{ \begin{array}{cl} (-)^{p-q}\binom{t-q}{t-p} & \text{if } q \leq p \\ 0 & \text{otherwise} \end{array} \right. \qquad (10)$$

Matrices N, V, and R. Putting everything together we have

$$R_k^s = \overline{D}_k^i P_i^p F_p^q C_q^r D_r^s$$

The (linear) map R_k^s is a polynomial in k of degree (at most) $t-1$. We can find its coefficients by rewriting

$$\begin{aligned} \overline{D}_k^i P_i^p &= P_k^p - P_{k-1}^p = (\tfrac{k}{n})^p - (\tfrac{k-1}{n})^p \qquad (11) \\ &= \sum_{l=1}^{p} k^{l-1}(-)^{p-l}\binom{p}{l-1}(\tfrac{1}{n})^p = V_k^l N_l^p \end{aligned}$$

where

$$V_k^l := k^{l-1}, \quad \text{and} \qquad (12)$$

$$N_l^p := \left\{ \begin{array}{cl} (-)^{p-l}\binom{p}{l-1}(\tfrac{1}{n})^p & l \leq p \\ 0 & \text{otherwise} \end{array} \right. \qquad (13)$$

4

Hence we get the alternative representation

$$\pi_k = R_k^s \alpha_s = V_k^l N_l^p F_p^q C_q^r D_r^s \alpha_s \qquad (14)$$

Injective. Matrices D, $\overline{D}$, and F are lower-triangular matrices with 1 in the diagonal, and hence are invertible (thus injective). C is diagonal and N upper triangular, both nowhere zero on the diagonal, hence invertible too. The first t rows of V map from a set of t coefficients $\boldsymbol{b}$ to the polynomial $p(x) = \sum_{l=1}^t b_l x^{l-1}$ evaluated at $x = 1, 2, 3, ..., t$. A degree $t-1$ polynomial like p is uniquely determined by t image points (see Appendix), hence V is injective. Similarly for P or exploit $P_k^l = k V_k^l (\frac{1}{n})^l$ (no summation). This proves that R is injective.

Matrix T. Combining the map from $\boldsymbol{a}$ to $\boldsymbol{\pi}$

$$\pi_k \equiv P(I = k) = \sum_{l=1}^t a_l k^{l-1} = V_k^l a_l \quad \text{i.e.} \quad \boldsymbol{\pi} = V\boldsymbol{a},$$

with $a_l = T_l^s \alpha_s$ we get

$$\pi_k = V_k^l T_l^s \alpha_s$$

Comparing this with (14) and using injectivity of V we see that

$$T_l^s = N_l^p F_p^q C_q^r D_r^s \qquad (15)$$

which is injective, since N, F, C, and D are invertible. $\qquad \square$

Discussion. Given a Polynomial Rank scheme it is possible and easy (using computer software) to find if it is equivalent to a probabilistic tournament (and get the corresponding parameters) by applying the inverse of T to $\boldsymbol{a}$. If the output satisfies the probability requirements $\boldsymbol{\alpha} \in \Delta_t$, then it is indeed a probabilistic tournament.

4 Map from Polynomial Ranking to Tournaments

We now derive explicit expressions for the really interesting converse of map T, which allows replacement of inefficient rank selections by equivalent efficient tournaments. From the last section we know that the inverse exists.

Theorem 5 (polynomial→tournament) *The function $\overline{T} : I\!R^t \to I\!R^t$, mapping polynomial coefficients $\boldsymbol{a}$ to tournament parameters $\boldsymbol{\alpha}$ is linear*

$$\alpha_s = \sum_{l=1}^t \overline{T}_s^l a_l, \quad i.e. \quad \boldsymbol{\alpha} = \overline{T}\boldsymbol{a}$$

where matrix $\overline{T} = T^{-1} = \overline{D}\,\overline{C}\,\overline{F}\,\overline{N}$, with $\overline{D}, \overline{C}, \overline{F}, \overline{N}$ defined in (7), (16), (17), (18). $\boldsymbol{a}$-polynomial ranking can be implemented as an $\boldsymbol{\alpha}$-tournament if and only if, $\boldsymbol{\alpha} = \overline{T}\boldsymbol{a} \in \Delta_t$.

Inverse matrices. In the following P and V respectively denote the upper $t \times t$ submatrix of P and V. The inverse matrices are as follows

$$\overline{C}_r^q := \delta_{r,q}/\binom{t}{r} \qquad (16)$$

$$\overline{F}_q^r := \binom{t-r}{t-q} \text{ if } r \le q \text{ and } 0 \text{ else} \qquad (17)$$

$$\overline{N}_p^l = \overline{P}_p^\iota D_\iota^\kappa V_\kappa^l \qquad (18)$$

$$\overline{P}_l^\kappa := n^l \overline{V}_l^\kappa \frac{1}{\kappa} \qquad \text{(no summation)} \qquad (19)$$

The inverse of the diagonal matrix C is obvious. The expression for $\overline{P}$ immediately follows from $P_\kappa^l = \kappa V_\kappa^l (\frac{1}{n})^l$ (no summation).

$F_p^q \overline{F}_q^r = 0$ for $r > p$ (since then either $r > q$ or $q > p$) and for $r \le p$ we have

$$F_p^q \overline{F}_q^r = \sum_{q=r}^p (-)^{p-q} \binom{t-q}{t-p} \binom{t-r}{t-q}$$
$$= \binom{t-r}{t-p} \sum_{q=r}^p \binom{p-r}{q-r} (-)^{p-q} = \delta_{p,r}$$

The first equality is by definition, the second equality is a simple reshuffling of factorials, and the last equality follows from the well-known binomial identity $\sum_{i=0}^m (-)^i \binom{m}{i} = 0$ for $m \ge 1$. This proves that $\overline{F}$ is the inverse of F.

Unfortunately we were not able to invert N directly, although N seems similar to (the transpose of) F. So we used relation (11) to invert N in (18). But now we need the inverse of P, which can be reduced by (19) to the inverse of V.

Inverse of V. The most difficult matrix to invert is V. This special Vandermonde matrix V can be written as a product of a lower L and upper-triangular matrix U, whose inverses are [Tur66]:

$$\overline{V}_l^\kappa := \overline{U}_l^s \overline{L}_s^\kappa$$

$$\overline{L}_s^\kappa := \frac{(-)^{s-\kappa}}{(s-\kappa)!(\kappa-1)!} \text{ for } s \ge \kappa \text{ and } 0 \text{ else}$$

$$\overline{U}_l^s := S_s^{(l)} = \text{Stirling numbers of the first kind}$$

The Stirling numbers $S_s^{(l)}$ numbers are defined as the coefficients of the polynomial $x(x-1)...(x-s+1)$, i.e. by

$$\sum_{l=0}^s S_s^{(l)} x^l = \frac{x!}{(x-s)!} \quad \text{and} \quad S_s^{(l)} = 0 \text{ for } l > s$$

There are many ways to compute $S_s^{(l)}$, e.g. recursively by $S_{s+1}^{(l)} = S_s^{(l-1)} - s S_s^{(l)}$ or directly [AS74, p.824]. For $r \ge \kappa$ we get

$$V_r^l \overline{V}_l^\kappa = \sum_{l=1}^t r^{l-1} \sum_{s=\kappa}^l \frac{S_s^{(l)} (-)^{s-\kappa}}{(s-\kappa)!(\kappa-1)!}$$

$$= \sum_{s=\kappa}^t \frac{[\sum_{l=1}^s r^{l-1} S_s^{(l)}](-)^{s-\kappa}}{(s-\kappa)!(\kappa-1)!}$$

$$= \sum_{s=\kappa}^t \frac{[(r-1)...(r-s+1)](-)^{s-\kappa}}{(s-\kappa)!(\kappa-1)!}$$

$$= \binom{r-1}{\kappa-1} \sum_{s=\kappa}^r \binom{r-\kappa}{r-s} (-)^{s-\kappa} = \delta_{\kappa,r}$$

The case $r<\kappa$ is similar. This shows that $\overline{V}$ is the inverse of (the first t rows of) V.

Linear ranking example. For $t=d+1=2$ we can compute the matrices by hand. This list of (reduced) matrices is a useful sanity check for the reader's own implementation:

$$F = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad \overline{F} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad H = \tfrac{1}{n^2}\begin{pmatrix} n-1 & 2n-4 & 3n-9 & \cdots & 0 \\ 1 & 4 & 9 & \cdots & n^2 \end{pmatrix}^{\top}$$

$$C = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \overline{C} = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad G = \tfrac{1}{n^2}\begin{pmatrix} n-1 & n-3 & \cdots & -1+n \\ 1 & 3 & \cdots & 2n-1 \end{pmatrix}^{\top}$$

$$N = \tfrac{1}{n^2}\begin{pmatrix} n-1 \\ 0 & 2 \end{pmatrix}, \qquad\qquad \overline{N} = \tfrac{n}{2}\begin{pmatrix} 2 & 1 \\ 0 & n \end{pmatrix}$$

$$P = \tfrac{1}{n^2}\begin{pmatrix} n & 2n & 3n & \cdots & n^2 \\ 1 & 4 & 9 & \cdots & n^2 \end{pmatrix}^{\top}, \qquad \overline{P} = \tfrac{n}{2}\begin{pmatrix} 4 & -1 \\ -2n & n \end{pmatrix}$$

$$V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & n \end{pmatrix}^{\top}, \qquad\qquad \overline{V} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$$

$$T = \tfrac{1}{n^2}\begin{pmatrix} 2n+1 & -1 \\ -2 & 2 \end{pmatrix}, \qquad\qquad \overline{T} = \tfrac{n}{4}\begin{pmatrix} 2 & 1 \\ 2 & 2n+1 \end{pmatrix}$$

$$\overline{U} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad \overline{L} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad R = \tfrac{1}{n^2}\begin{pmatrix} 2n-1 & 2n-3 & \cdots & 1 \\ 1 & 3 & \cdots & 2n-1 \end{pmatrix}^{\top}$$

We see that $\boldsymbol{\pi}=R\boldsymbol{\alpha}$ and $\boldsymbol{a}=T\boldsymbol{\alpha}$ coincide with (4), as they should.

Computational complexity. Together this allows us to compute $\boldsymbol{\alpha}$ from $\boldsymbol{a}$ and vice versa in time $O(t^2)$ and $\boldsymbol{\pi}$ from $\boldsymbol{\alpha}$ in time $O(nt)$. Once $\boldsymbol{\alpha}$ is known, tournament selection needs only time $O(t)$ per winner selection.

5 What Polynomial Selection Schemes are Tournaments?

Theorem 5 does not give us conditions under which the resulting tournament parameters $\boldsymbol{\alpha}=\overline{T}\boldsymbol{a}$ are valid. We look for such conditions so that we can reliably change/create tournament schemes in the more understandable set of polynomial rank schemes. Without these conditions there can be no guarantee that whatever created would be a probabilistic tournament.

Range of linear ranking. Let us first consider the case of linear ranking ($d=1$),

$$P(I = k) = a_1 + a_2 k$$

We want to find the range of a_1 and a_2 for which this is a proper probability distribution in Δ_n. The sum-constraint leads to

$$1 = \sum_{k=1}^{n} P(I = k) = a_1 n + a_2 \tfrac{1}{2} n(n + 1)$$

$$\implies \quad a_1 = \tfrac{1}{n}[1 - \tfrac{1}{2}a_2(n^2 + n)] \tag{20}$$

Next are the positivity constraints $P(I = k) \geq 0 \;\forall k$. A linear function is ≥ 0 if and only if it is ≥ 0 at its ends, i.e. $P(I=1)\geq 0$ and $P(I=n)\geq 0$. Inserting (20) into these constraints yields:

$$P(I = 1) \;\equiv\; a_1 + a_2 \;\geq 0 \quad\Longleftrightarrow\quad a_2 \leq \tfrac{2}{n^2-n}$$

$$P(I = n) \;\equiv\; a_1 + a_2 n \geq 0 \quad\Longleftrightarrow\quad a_2 \geq -\tfrac{2}{n^2-n}$$

So the possible linear rank schemes are those with

$$|a_2| \leq \tfrac{2}{n^2-n} \quad \text{and} \quad a_1 \text{ satisfying (20)} \tag{21}$$

Range of tournament size 2. The example (4) shows that size $t = 2$ probabilistic $\boldsymbol{\alpha}$-tournaments have $a_2 = 2(\alpha_2 - \alpha_1)/n^2$. Since $\boldsymbol{\alpha} \in \Delta_2$, a_2 has range $-\tfrac{2}{n^2}...\tfrac{2}{n^2}$. As it should be, this is a subset of the possible linear rank schemes. Hence the linear rankings that are probabilistic tournaments are those with

$$|a_2| \;\leq\; \tfrac{2}{n^2} \quad \text{and} \quad a_1 \text{ given by (20)} \tag{22}$$

This is slightly narrower than $|a_2| \leq \tfrac{2}{n^2-n}$, i.e. there are some rankings that are not probabilistic tournaments. On the other hand, $\tfrac{2}{n^2-n}/\tfrac{2}{n^2}$ tends to 1 as n grows, hence for n large (e.g. about 100) nearly all linear rankings can be translated into probabilistic tournaments. The coverage is good enough for all practical purposes.

The general case. A probabilistic selection scheme is completely determined by $\boldsymbol{\pi}$, different $\boldsymbol{\pi}$ correspond to different selection schemes, and every $\boldsymbol{\pi} \in \Delta_n$ is a valid selection scheme. Hence, Δ_n is the set of all possible probabilistic selection schemes. The set of (valid) size t tournament schemes is

$$R\Delta_t \;:=\; \{\boldsymbol{\pi} = R\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \Delta_t\} \;\subset\; \Delta_n$$

Since R is injective, this is a $t-1$ dimensional irregular simplex embedded in the $n-1$ dimensional simplex Δ_n.

The set of (incl. invalid) degree (up to) $t-1$ polynomial ranking schemes is

$$V{I\!\!R}^t \;:=\; \{\boldsymbol{\pi} = V\boldsymbol{a} : \boldsymbol{a} \in {I\!\!R}^t\} \;\not\subseteq\; \Delta_n$$

This is a t-dimensional hyperplane. Only $\boldsymbol{\pi}$ in Δ_n are valid, hence $V{I\!\!R}^t\cap\Delta_n$ is the set of (valid) polynomial ranking schemes. The intersection of a simplex with a plane gives a closed, bounded, convex polytope, in our case of dimension $t-1$. The Krein-Milman Theorem [Edw65, p.707] says that for a closed, bounded, convex subset A of ${I\!\!R}^t$ with a finite number of extreme points (=corners), A is the convex hull of the extreme points of A. Hence the extreme points of $V{I\!\!R}^t\cap\Delta_n$ completely characterize/define the set.

If/since we are not concerned with the covering of $V{I\!\!R}^t\cap\Delta_n$ in Δ_n itself, we can study the covering in the lower-dimensional polynomial coefficient space ${I\!\!R}^t$. The set=polytope of all polynomial coefficients $\boldsymbol{a}$ that lead to valid selection probabilities is

$$\overline{V}\Delta_n \;:=\; \{\boldsymbol{a} \in {I\!\!R}^t : V\boldsymbol{a} \in \Delta_n\}$$

while the set=simplex of coefficients reachable by tournaments is

$$T\Delta_t \;:=\; \{\boldsymbol{a} = T\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \Delta_t\} \;\subset\; \overline{V}\Delta_n$$

These sets are the images of $V{I\!\!R}^t\cap\Delta_n$ and the simplex Δ_t under $\overline{V}$ and T respectively. These maps are injective (Section 4) so $\overline{V}\Delta_n$ and $T\Delta_t$ are completely determined by their extreme points. The extreme points of Δ_t are just the conventional ${I\!\!R}^t$ basis vectors $\boldsymbol{e}_s$, so $T\Delta_t$ is the convex

hull of $\{T(\boldsymbol{e}_s) : s = 1,...,t\}$. The polytope $V\mathbb{R}^t \cap \Delta_n$ can be quite complex, and finding the extreme points daunting. This is essentially what we did for the $t = 2$ case in the above paragraphs.

We estimated the proportion of degree $t-1$ polynomials covered by $T\Delta_t$ for various t using a Monte-Carlo algorithm[1] (Table 1). It shows that for $n \geq 100$, practically all linear rank schemes are probabilistic tournaments.

Nothing concrete can be concluded about the coverage for $t = 4,5$. Table 1 only suggests that the number of $t-1$ degree polynomials equivalent to t-sized tournaments decreases as t increases.

Tournament size 3. In the $t = 3$ case we can extend our knowledge by finding $\overline{V}\Delta_n$ graphically. The restriction $\sum_{k=1}^n \pi_k = 1$ means that the k^2 coefficient a_3, is completely determined by a_1 and a_2.

$$a_3 = \frac{1}{\sum_{k=1}^n k^2} \left(1 - na_1 - \tfrac{1}{2}n(n+1)a_2\right) \qquad (23)$$

Hence $\overline{V}\Delta_n$ is a 2 dimensional hyperplane. $P(I = k) \geq 0$ for each k defines a set of halfspaces: $\{(a_1,a_2,a_3) : a_1 + a_2k + a_3k^2 \geq 0\}$; $\overline{V}\Delta_n$ is their intersection (over k) restricted to the plane given by (23).

$T\Delta_3$ is simply a filled triangle with corners $\{T(\boldsymbol{e}_s) : s = 1,2,3\}$. Comparison with $\overline{V}\Delta_n$ (Figures 2, 3 and 4) suggests that the coverage of $T\Delta_3$ is stable for $n \to \infty$. Hence for large populations about a third of the quadratic polynomials can be written as size-3 probabilistic tournaments.

In practice, selection schemes with probability monotonically increasing with fitness are used. So not the whole of $\overline{V}\Delta_n$ is interesting, but only the subset of monotonically increasing or possibly decreasing probabilities on $\{1,2,...,n\}$ (light grey in figures 2, 3 and 4). The remainder of $\overline{V}\Delta_n$ is composed of schemes that favour the middle ranks or both high and low ranked individuals (dark grey).

Any polynomial scheme $P(I = k) = a_1 + a_2k + \frac{1}{\sum_{i=1}^n i^2}\left(1 - na_1 - \tfrac{1}{2}n(n+1)a_2\right)k^2$ is a parabola[2], so it is symmetric about it's stationary point, $x_{st.pt.}$. Hence $P(I = k)$ is monotonic on $\{1,2,...,n\}$ if and only if $x_{st.pt.}$ lies outside the interval $(1 + \tfrac{1}{2}, n - \tfrac{1}{2})$.

i.e.

$$x_{st.pt.} = \frac{-a_2}{1 - na_1 - \tfrac{1}{2}n(n+1)a_2}\left(\sum_{i=1}^n i^2\right) \leq 1 + \frac{1}{2}$$

OR

$$x_{st.pt.} \geq n - \frac{1}{2}$$

Figure 4 suggests that these regions of usefulness effectively lie entirely in $T\Delta_3$ for $n \geq 300$. Hence for sufficiently large n the most useful degree 2 polynomial schemes are perfectly reproduced by some probabilistic tournament.

An example of a less applicable selection scheme is the polynomial given by $a_1 = 0.01$ and $a_2 = -1 \times 10^{-4}$ (which

---

[1] The $t = 2$ case was calculated directly from (21) and (22)

[2] We temporarily consider k to range over the real line

	n = 4	**n = 10**	**n = 20**	**n = 100**	**n = 300**
t = 2	0.7500	0.9000	0.9500	0.9900	0.9967
t = 3	0.270		0.348	0.342	0.332
t = 4		0.12	0.15	0.16	
t = 5		0.02			

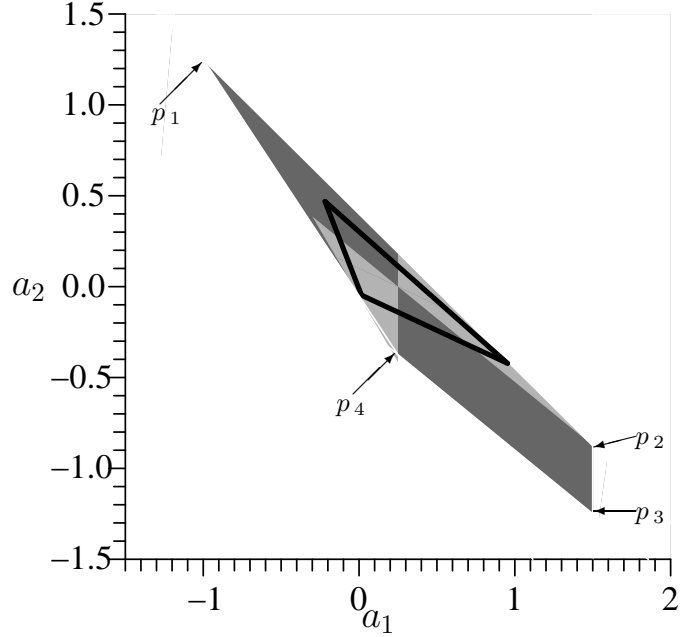Table 1: fraction of possible $t-1$ degree polynomials that car

Figure 2: $[n = 4, t = 3]$ *The shaded region is the set of possible polynomials, whilst the light grey area is the set of the most useful polynomials. The triangle is the boundary of the set that can be written as $t = 3$ tournaments. At p_1: $a_3 \doteq -0.246$. At p_2: $a_3 \doteq 0.159$. At p_3: $a_3 \doteq 0.236$. At p_4: $a_3 \doteq 0.023$.*

lies in the dark grey region). It favours both high ranks and low ranks (Figure 5) and any algorithm using this scheme will spend half of the time searching in the wrong place. However it is still usable (like in fitness uniform selection [HL06]).

The points p_1, p_2 ... are extreme points of $\overline{V}\Delta_n$. They indicate that the range of a_3 values is significantly smaller than the range of a_2 (which in turn has a smaller range than a_1).

$\overline{V}\Delta_n$ being the intersection of a finite number of halfspaces and planes means its boundary is actually a series of straight lines. $\overline{V}\Delta_n$ appears curved in figures 3 and 4 simply due to the many halfspaces that are involved.

6 Discussion/Conclusions

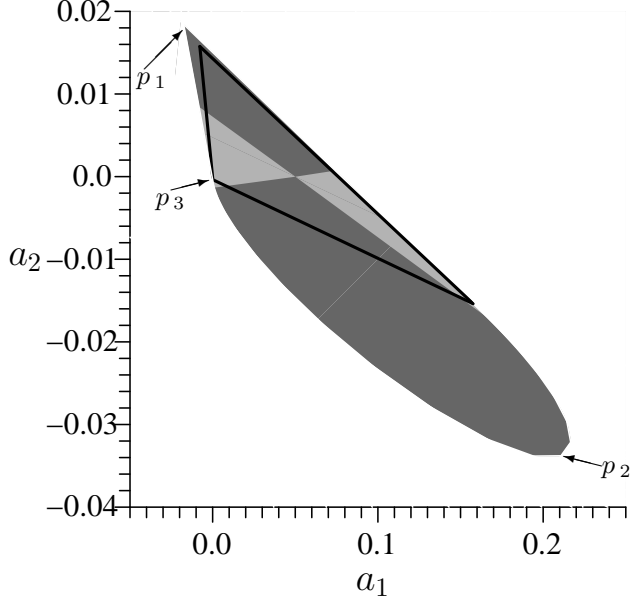**Rank ties.** Individuals with the same fitness lead to ties in the ranking. If we break ties (arbitrarily but consis-

Figure 3: $[n=20, t=3]$ *The shaded region is the set of possible polynomials, whilst the light grey area is the set of the most useful polynomials. The triangle is the boundary of the set that can be written as $t=3$ tournaments. At p_1: $a_3 \doteq -8.56 \times 10^{-4}$. At p_2: $a_3 \doteq 1.08 \times 10^{-3}$. At p_3: $a_3 \doteq 1.91 \times 10^{-4}$.*
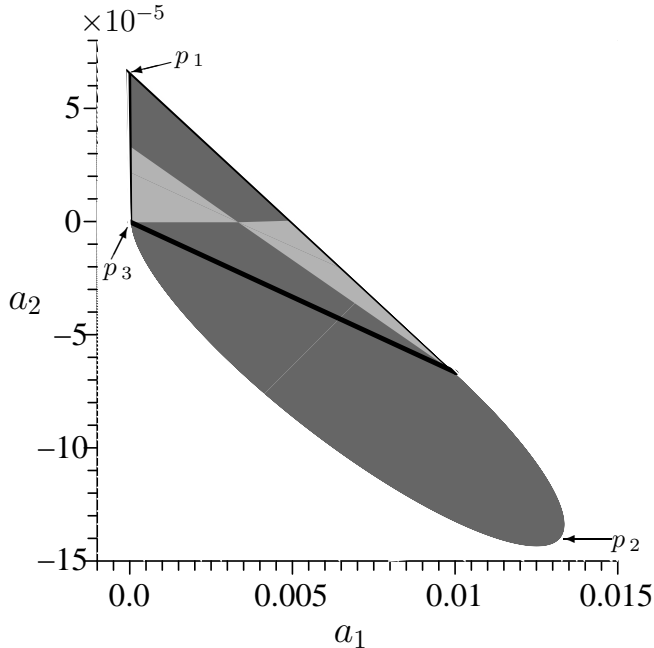

Figure 4: $[n=300, t=3]$ *The shaded region is the set of possible polynomials, whilst the light grey area is the set of the most useful polynomials. The triangle is the boundary of the set that can be written as $t=3$ tournaments. At p_1: $a_3 \doteq -2.18 \times 10^{-7}$. At p_2: $a_3 \doteq 3.53 \times 10^{-7}$. At p_3: $a_3 \doteq 1.21 \times 10^{-7}$.*
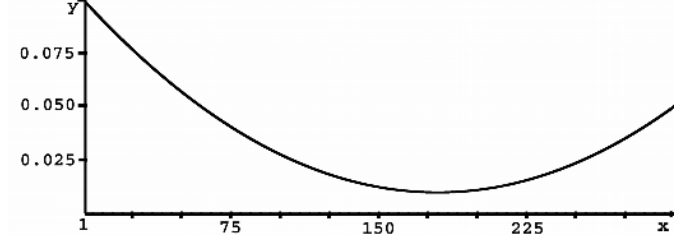

Figure 5: [n=300] The polynomial $y = 0.01 - 10^{-4}x + 2.781 \times 10^{-7}x^2$. This is an example of a usable quadratic polynomial that is not equivalent to a probabilistic tournament.

tently), our theorems still apply. The disadvantage is that the selection probability for two individuals with the same fitness may not be the same. We can fix this problem by breaking ties (uniformly) at random. For instance, given a population of 3 individuals with two of them having the same fitness, this results in effective selection probabilities $\pi_1^{eff} = \pi_1$ and $\pi_2^{eff} = \pi_3^{eff} = \frac{1}{2}(\pi_2 + \pi_3)$.

Further work. Investigation of the set of possible polynomials with degree $d \geq 3$ will be helpful for those applications requiring higher selective pressures. Furthermore, finding the proportion that are equivalent to probabilistic tournaments may provide a reliable method for making high-degree polynomial rank schemes more efficient.

Tournaments of size $t \ll n$ are significantly faster than ranking schemes, so it would be beneficial to obtain a thorough understanding of how many polynomial rank schemes are equivalent to $t > d+1$ sized probabilistic tournaments.

Conclusion. We have found a strong connection between polynomial ranking and probabilistic tournament selection.

We derived an explicit operator (15) that maps any probabilistic tournament to its equivalent polynomial ranking scheme, which is unique and always exists. Polynomial rank schemes thus encompass linear ranking and deterministic (normal) tournament selection, leaving designers with one less selection method (but more parameters) to worry about.

Unfortunately, turning polynomial rank schemes into equivalent probabilistic tournaments is not so straightforward. Only about a third of the possible quadratic polynomials can be written as size-3 probabilistic tournaments.

However, nearly all linear rank schemes have an equivalent size-2 probabilistic tournament. Hence nearly all can be made faster by simply rewriting the scheme as a probabilistic tournament.

Furthermore, almost all the practical quadratic polynomials are equivalent to some $t=2$ tournament. This is a good indication for the investigation of $t > 3$.

A Appendix

Uniqueness of a polynomial given t image points.
Let $\boldsymbol{\pi} \in I\!\!R^t$ be the vector of t image points $\pi_\kappa = p(x_\kappa)$ for some $x_1,...,x_t$ of a polynomial $p(x) = \sum_{l=1}^{t} a_l x^{l-1}$ with coefficient vector $\boldsymbol{a} \in I\!\!R^t$. In particular we have

$$\pi_\kappa \;=\; p(x_\kappa) \;=\; \sum_{l=1}^{t} a_l V_\kappa^l, \quad \text{where} \quad V_\kappa^l \;=\; x_\kappa^{l-1}$$

If matrix V is invertible, the polynomial (coefficients) would be uniquely defined by $\boldsymbol{a} = V^{-1}\boldsymbol{\pi}$, which is what we set out to prove. We now show that V is invertible. Define the t polynomials of degree $t-1$

$$p_s(x) \;=\; \prod_{\substack{r=1 \\ r \neq s}}^{t} \frac{x - x_r}{x_s - x_r} \;=\; \sum_{l=1}^{t} A_l^s x^{l-1}$$

Expanding the product in the numerator defines the coefficients A_l^s. On x_κ we get

$$\delta_{s\kappa} = p_s(x_\kappa) = \sum_{l=1}^{n} V_\kappa^l A_l^s$$

hence A is the inverse of V. By explicitly expanding $\prod(x - x_r)$ one can get an explicit expression for A_l^s, which is unfortunately pretty useless.

References

[AS74]   M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover publications, 1974.

[Bäc94]   T. Bäck. Selective pressure in evolutionary algorithms: a characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, volume 1, pages 57–62, Orlando, FL, USA, 1994. IEEE World Congress on Computational Intelligence.

[BT95]   T. Blickle and L. Thiele. A comparison of selection schemes used in genetic algorithms. TIK-Report 11, TIK Institut fur Technische Informatik und Kommunikationsnetze, Computer Engineering and Networks Laboratory, ETH, Swiss Federal Institute of Technology, Gloriastrasse 35, 8092 Zurich, Switzerland, 1995.

[Edw65]   R. E. Edwards. *Functional Analysis: Theory and Applications*. Holt, Rinehart and Winston, Inc, USA, 1965.

[ES03]   A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Springer, 2003.

[Fog88]   D. B. Fogel. An evolutionary approach to the travelling salesman problem. *Biological Cybernetics*, 6(2):139–144, 1988.

[GD91]   D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In G. J. E. Rawlings, editor, *Foundations of genetic algorithms*, pages 69–93. Morgan Kaufmann, San Mateo, 1991.

[Gol89]   D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Mass., 1989.

[HL06]   M. Hutter and S. Legg. Fitness uniform optimization. *IEEE Transactions on Evolutionary Computation*, 10:568–589, 2006.

[Hut91]   M. Hutter. Implementierung eines Klassifizierungs-Systems. Master's thesis, Theoretische Informatik, TU München, 1991. 72 pages with C listing, in German, http://www.idsia.ch/~marcus/ai/pcfs.htm.

[Tur66]   L. R. Turner. Inverse of the Vandermonde matrix with applications. Technical Report NASA TN D-3547, Lewis Research Center, Cleveland, Ohio, 1966.

[WC02]   W. Wieczorek and Z.J. Czech. Selection schemes in evolutionary algorithms. In *Intelligent Information Systems 2002, Proceedings of the IIS'2002 Symposium, Sopot, Poland, June 3-6, 2002*, Advances in Soft Computing, pages 185–194. Physica-Verlag, 2002.